DOCUMENT RESUME

ED 327 558 TM 015 965

AUTHOR Ingvarson, Lawrence
TITLE Enhancing Professional Skill and Accountability in the Assessment of Student Learning.
PUB DATE Mar 90
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Accountability; Conferences; Foreign Countries; High Schools; *Learning; Models; Principals; Secondary School Students; *Secondary School Teachers; *Skill Development; *Student Evaluation
IDENTIFIERS Australia; *Consensus Moderation

ABSTRACT

This paper presents a procedure for making high school teachers' assessments comparable across schools for graduation and college admission purposes. Findings are presented of an evaluation of the procedure conducted in Victoria (Australia) between 1981 and 1984. This procedure, called "consensus moderation," compares teachers' assessments with clear criteria for the assessment tasks, criteria that were developed by groups of teachers. The procedure requires teachers from different schools to meet in groups of 10 to 12 at least 3 times per year with the aim of aligning grades with quality of work. Members of the Australian evaluation team directly observed the work of 6 of 30 consensus moderation groups for the first 4 years of their operation. The team also participated in central meetings and received the results of surveys of geography teachers (n=100), principals, and students. Teachers, and the vast majority of students and principals, were generally confident that the procedure was fair. The process appeared to support teachers by giving weight to their assessments and to reinforce and clarify the links between the models of assessment that teachers chose and their curriculum objectives. Five tables present teacher survey findings. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED327558

Enhancing Professional Skill and Accountability in the Assessment of Student Learning

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

LAWRENCE INGVARSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Lawrence Ingvarson
Monash University
Clayton
Australia, 3126

Paper presented at the 1990 Annual Meeting of the American Educational Research Association, Boston

TM015965

Enhancing Professional Skill and Accountability in the Assessment of Student Learning

Lawrence Ingvarson
Monash University

Prevailing practice in student assessment in the U.S.A. undermines teacher professionalism, according to some commentators (Calfee & Hierbert, 1988, Stake, 1989). The ability to devise appropriate methods for gathering information about student learning is a central skill of the professional teacher. Without it, it is hard to see how teachers can have a sound basis for evaluating their own practice. However, professional development in this area will be undermined if more weight is placed on the outcomes of standardised tests than teachers' assessments of their students' achievements. So, it seems odd that, in the recent widespread discussion about "empowering" teachers professionally and "restructuring" schools in the U.S.A., few reform proposals address the issue of how to shift the balance of status and influence in favour of teacher judgment over standardised tests. In fact, since 1983 the importance of standardised testing of student achievement has escalated. (Stake, 1988)

The focus of this paper is a procedure for making high school teachers' assessments comparable across schools for graduation and college admission purposes. Introduced in the 1981, the procedure is called "consensus moderation", and it has had impressive side effects on the professional development and accountability of teachers. The setting is the state of Victoria in Australia. The paper presents part of the findings of an evaluation of these procedures which was conducted from 1981 to 1984 together with a follow up study in 1989.

Background

Some background to this paper seems to be in order. In 1989, I was fortunate to be able to spend eight months study leave in the U.S. working in the OERI funded Center for Research on the Context of Secondary School Teaching at Stanford University, directed by Milbrey McLaughlin. A key aspect of the Center's brief is to examine the contextual factors and workplace conditions that shape high school teacher's attitudes to their work and their ability to carry out their work as well as they would wish; in particular, the level of "engagement" of teachers and students in the work and life of the school, the interdependence between these, and the way in which they are shaped by the school setting.

As part of the research team, I conducted interviews with about forty teachers in several Californian high schools. These articulate, experienced and well qualified teachers were highly committed, particularly to the personal welfare of their students. At the same time, relative to Victorian teachers, many conveyed a sense of impotence, or at least un-involvement, in relation to wider contextual influences that appeared to me to have an important bearing on their attitudes to their work, on what they taught and how they taught, and, in particular, the status given to their assessments of what their students had learned.

Some examples of contextual differences would include the following:

> There are no externally mandated tests in Victoria. A coalition of teacher and parent organisations has successfully thwarted several attempts by state governments to introduce any form of state-wide testing.

The state education authority does not involve itself in the selection of textbooks. That is considered to be a matter for teacher judgment. Consequently, book companies visit schools to hawk their wares directly to teachers.

Participatory decision-making in the workplace, with respect to administrative and curriculum matters, is built into the union-employer agreement in Victoria. For example, "Administrative Committees" in each school, consisting of staff representatives and the principal, make decisions each year about the allocation amongst the staff of "positions of responsibility" which carry extra salary loadings.

Teachers and parents on school-site councils sit on committees to select principals. The same councils have the responsibility for establishing and reviewing the school's educational policy (which would include the school's policy on assessment.), within very broad guidelines laid down by the State Government.

(None of the above, however, should be taken to imply that teacher morale and job satisfaction is any higher than in the US. Judging by recent surveys the situation is about equally bad in both countries.)

What interested me in particular was the value that the Californian high school teachers felt was placed on their assessments of what their students had learned; and, by implication, what they had taught. A regular theme in the teacher interviews went something like, "I could teach what ever I liked. No one really knows, or cares". Did teachers develop their own methods for assessment, or did they use they use pre-designed tests put out with the textbooks? One biology teacher expressed considerable frustration with the low level, and irrelevance to her teaching, of the content being tested in one of the text book tests she was using. However, the idea of developing her own modes of assessment to better cover the range of objectives she was teaching toward, such as practical work or fieldwork, or even multiple choice items that went beyond testing for recall, was not an option she considered feasible, given the pressures under which she worked.

Although teacher grades were taken into account for college entrance purposes, there was no moderation process for ensuring that there was some comparability in teachers' standards from school to school. Was the status and credibility of teacher grades, I wondered, lower as a result, in comparison, say, with Scholastic Aptitude Test scores? Did teachers feel as accountable for the SAT scores their students gained as they did for their own gradings? Which assessments really counted? California has a comprehensive program of state-wide testing, the California Assessment Program, in which all public school students in grades 3, 6, 8, and 12 are tested annually. Did teachers feel accountable for the results their students gained in these standardised tests? If the status and the credibility of teacher assessments was weakened because there was no mechanism for ensuring comparability and professional accountability, did this have an adverse impact on teachers' engagement and job satisfaction?

So far as Advanced Placement classes were concerned, the fieldwork left no doubt that teaching at that level was of was high status, and that teachers who had AP classes invested a disproportionate amount of their energy in that part of their work. Teachers felt accountable for the results their students gained. However, AP programs prepared only a small proportion of students for an examination set and graded by an external College Board. Teachers appeared to have little opportunity to play a part in determining the content of the course, or the modes of assessment.

At the same time I became very interested in the wider policy debate about the "restructuring" of teachers' work. A striking feature of these reforms, to an outsider, was the way in which they were aimed directly at raising the professional status of teaching, almost as an end in itself, rather than, for example, a outcome of a wider educational reform of curriculum or pedagogy whose need had been identified. Professionalisation seemed to be the reform itself. There was much talk of a crisis in teaching, and the major vehicle for overcoming the crisis was the "professionalisation" of teaching, which was equated with the "empowering" of teachers. But empowerment for what? The answer in reports such as *A Nation Prepared* (1986) appears to be, in order to do better at what we are currently trying to do, to raise standards, "to graduate the vast majority of . . .students with achievement levels long thought possible for only the privileged few" (Carnegie Forum On Education And The Economy, p. 21) Empowerment in the stronger sense of authorising, delegating or entrusting teachers to question what they are currently being asked to do, or to participate more directly in *curriculum* policy and reform at school, district and state level, does not receive the same emphasis.

What kinds of situations, experiences or responsibilities are empowering for teachers as teachers? By definition, a crucial characteristic is that such situations *authorise* teachers to exercise professional responsibility. They support, rather than supplant, the exercise of teacher judgment. Judgment can only be developed by using it. They validate and reinforce, rather than replace, professional discretion. As a corollary, they call for different forms of accountability from the surrogate standardised test score: forms which are accounts or descriptions of how professional judgment has been exercised; activities which open up teachers work to each other. In the context of this paper, this will mean accounts which teachers provide to each other describing the methods they have used for assessing student progress and achievement, which thereby enhance their professional acccountability (Darling-Hammond, 1989)

I was aware of the responsibility that high school teachers in Victoria had argued for and taken upon themselves in restructuring the curriculum and assessment methods used in public examinations; and that this had had strong professionalising and empowering effects. This was particularly true for the process of moderation, even though they were a side effect to the main task. When I described the moderation methods that were used in Victoria to American friends, it usually evoked strong interest. So I wondered if there might be a lesson here.

So called "second wave" proposals for the reform of schoolwork in the US give emphasis to empowering teachers rather than managing them, and to expanding the scope of their responsibilities (Johnson, 1989). It is necessary therefore to identify specific areas into which the zone of professional responsibility can be extended. This paper is concerned with one that doesn't appear to have received much recognition in the US reform literature, the assessment of learning. Moderation pulls teachers together from different schools in order to carry out a task they see as important and relevant. In the process, it creates conditions for professional interaction which open up teachers' practice to each other. As well as increasing the sharing of ideas, these conditions inevitably lead to greater professional accountability (Darling-Hammond, 1989).

Origins of Moderation in Victoria

Until the 1970s no use was made in Victoria of teachers' grades of work that students completed at school, unlike the U.S.. The only basis for selecting students was their performance on State-wide subject-based examinations that students sat at the end of year 12. Students typically took five subjects including English which was compulsory. No

equivalent to the U.S. Scholastic Aptitude Test existed. Neither were student achievements in non-academic areas taken into account. Examinations were set and graded externally to the school by an independent "examinations board" dominated by the interests of the universities. The typical mode of assessment in most subjects was a three-hour essay-type test taken on the same day by all students near the end of the school year.

The use of teachers' assessments is now much more common in Australia, but a condition of their introduction was the development of procedures to ensure that teachers' standards and marks (grades) were comparable from school to school and teacher to teacher. This paper is based on an evaluation of one procedure designed to achieve this purpose called "consensus moderation". The particular version of it that is reported on here was developed largely by teachers in the state of Victoria in the late 1970s.

During the 1970s teachers' organisations in Victoria, including unions, played a leading role in curriculum reform. Control of the subject committees that designed courses and set the examinations was shifting from university academics to teachers. Innovative teachers were developing approaches to teaching their subjects as fields of enquiry rather than bodies of facts. But if most teachers were to take these reforms seriously and implement them in earlier high school years it was vital that the modes of assessment at the end of high school reflect these broader objectives. This is because examination marks are "high stakes". They form the sole basis for selection into University for school leavers. Not only is there a predictable relationship between what teachers emphasise and what the examinations examine; teachers also feel responsible for, and are held accountable for, the marks their students receive. This is especially so when, as in the case on which this paper is based, the forms of student assessment that "count" are an integral component of the curriculum and course design.

Teachers in several subject areas, such as Geography, Biology and English, wanted equal weight to be given to the external examination results and teachers' assessments in the final mark. They argued that school-based modes of assessment would provide more opportunity to teach toward objectives and skills whose attainment could not be assessed in a three hour examination situation. It would also give them greater flexibility to adapt the course guidelines to local conditions. As one of the teachers involved at the time put it

> We all believed that teachers were a very competent group of people - just as competent to make judgments about their students' work as a three hour examination. So we weren't just fighting for some idealistic principal. We actually believed that what we were after was both good for teachers and good for students. There was a real commitment to what we were doing.
>
> We knew there were excellent things happening in classrooms for which students got no credit at all. So we thought we ought to have information about a variety of ways in which students work - not only because that's a good teaching strategy, but because the kinds of careers that students were going into with this subject - town planning, etc. - continued the use of the skills they were learning at a professional level. So we thought there ought to be some sort of link between the real world of work that geographers did and what students did in school.

Others, such as university and business groups, wanted to retain what they saw as a credible and equitable basis for ranking and selecting students. They set a priority on common standards for comparing the quality of student achievement across schools. They made it clear that school-based assessments would only be acceptable if some method could

be found for ensuring that standards and grades were comparable for schools across the state.

The Consensus Moderation Process

Three methods were available to achieve this purpose: statistical moderation, moderation by visitation, and consensus moderation. Most subjects (there were 54 in all) with large student numbers were required to use statistical moderation, mainly because it was cheaper, politically "safe", and thought to be administratively simpler. English teachers were particularly disappointed that they did not win the case they had fought so hard for to move to consensus moderation, as so much of the development in that subject was based on new ideas about writing which emphasised the importance of audience, context and purpose in communication.

Statistical moderation was a straightforward process in which teachers sent in their grades to the examination authority and these were adjusted according to the performance of students from that school on the external examination. These school-based assessments usually count toward 30% of the final mark. Although there were guidelines about the forms of assessment that teachers should use, there was no real check on how teachers had arrived at their grades. This system, therefore, operates on the assumption that the work that teachers set for the purposes of school-based assessment tests the same skills and abilities as the external examination.

In moderation by visitation a person or a panel of persons visit the schools and review samples of student work, and in consultation with the teachers make adjustments if necessary.

Consensus moderation attempts to achieve comparability by comparing teachers' assessments with clear criteria for the assessment tasks, and with the assessments made by other teachers. In a year-long process, groups of teachers meet to establish mutually accepted conditions and criteria to be applied by teachers in making assessments, and in reviewing or "moderating" these assessments at the end of the year.

Geography teachers developed and gained approval for their own version of consensus moderation, as did eight other subjects with smaller numbers of students, such as art and music. According to a geography teacher who was involved

> They probably thought they could sacrifice Geography - if that went to the wall - lower status, smaller numbers. But they couldn't afford to take the same risk with English even though the teachers wanted it.

In brief, the procedures for moderation they developed require Year 12 teachers from different schools to meet in groups of about 10-12 at least three times per year. Each group is led by a convenor appointed by the examinations authority (usually a teacher from another district). The convenor's task is to chair meetings and ensure that the group works within the guidelines laid down by the state agency responsible. (Now called the Victorian Curriculum and Assessment Board.)

The term "consensus" is an indication that, in working towards comparability of assessments, the group takes responsibility for its decisions, and that decisions will be a result of agreement reached through discussion rather than voting procedures. But, if an impasse is reached in discussion the guidelines stipulate that the Convenor's decision will prevail.

At the first meeting each year, the moderation group members work toward a common understanding of the objectives and criteria for the assessment tasks. Each teacher must nominate the assessment methods they intend to use and show how they are related to the concepts and skills in the course outline. In Geography, two or three have to be chosen from a set of six which includes: a Fieldwork report, a Practical Work report, an Extended Essay, a teacher determined Item Bank, an Individual Research Project, and an Action Research Project. Every student is also required to carry out an Independent Research Project for another section of the course. By the mid-year meeting group members have usually circulated to each other for approval, details of the assessment tasks they are giving their students and the topics their students have chosen for their research projects.

Teachers are required to bring all the work that their students have done on the assessment tasks to the end of year meeting. The groups review (i.e. "moderate") samples of work selected randomly from each school by the convenor. Each piece of work is usually reviewed by at least three teachers. The aim is to bring into alignment the grades that had been given for work of similar quality from school to school. If the marks of the sample work from a school are adjusted, as a result of the review, then marks for all students from that school may be adjusted up or down.

This is a highly condensed summary of the procedures used. Although there are guidelines laid down for groups to follow, each has developed their own idiosyncratic ways of complying with them. An outline of the procedure used by one group on the final day can be found in Appendix 1.

At the end of the day the group chooses a sample of the whole group's work that is representative of their standards. This is taken by the convenor to a state-wide meeting of all convenors where much the same process of review is repeated. In a similar fashion, if the marks of the group's sample are adjusted as a result of the review, then marks for all students from that group may be raised or lowered.

Consensus moderation began in 1981, accompanied by considerable scepticism and controversy. There were doubts about the competence of teachers to make these judgments; whether they could make them objectively - or whether they would be subject to the pressure of their groups. There were suspicions that "consensus" would gravitate to "you scratch my back and I'll scratch yours". There were also concerns that the process would be disruptive and costly, as hundreds of teachers needed to be released from teaching and substitute teachers employed for two days each year.

Method

The evaluators adopted a responsive approach (Stake, 1976). Interviews and discussions with parties involved in or affected by the new procedures indicated that the evaluation should address questions and issues such as the following:

(a) Is consensus moderation fair to students, compared with the external examination?

(b) How reliable or consistent are teachers' assessments of each other's work? Are they as trustworthy or objective as external examinations?

(c) Does school-based assessment have a detrimental effect on student-teacher relationships? Does it increase workload unreasonably?

(d) Will consensus moderation groups actually moderate and adjust marks when necessary, or will they baulk when the going gets rough?

(e) What effect is the new procedure having on the quality of student work?

(f) How do moderation groups actually go about their task? What procedures facilitate or hinder achievement of the task?

Many more issues and unintended consequences emerged in the course of the evaluation. There has been a persistent tension between the desire for autonomy by groups and the need to have specific criteria and guidelines for the assessment items. The meaning of comparability itself underwent change, as did the term "quality" in talking about students' geographical thinking. Moderation came to be seen as a year-long process, not something done at the end-of-year meeting alone. Professional development of teachers involved was a powerful side-effect.

Data Sources

Members of the evaluation team directly observed the work of six of the thirty consensus moderation groups for the first four years of their operation, keeping records of the way each group made its decisions and carried out its tasks. Case studies developed from these records were used for formative evaluation purposes. They will not be reported in this paper.

The team also participated in the central meetings of the convenors, and the Geography Subject Committee which set policy. Each convenor was required to submit a detailed annual report of their meetings and the procedures they had used. These reports were also a valuable source of data.

All Geography teachers completed questionnaires at the end of each year for the first four years (1981-84). The same instrument was used once more at the end of the 1989 school year. A survey of principals was conducted at the end of the third year. Students were surveyed for the first two years. These are the main source of data used in this paper.

Various records were made available by the examining authority, such as student marks before and after moderation for each school and the marks that students gained on the external examination. Information about costs was also gathered.

Convenors were crucial people in the whole process and we conducted several group and individual interviews with them about their experiences.

Results

It was fortunate that this study was conducted over a period of four years. At the end of the first year the balance of evidence in favour of the new procedures was precarious. The difficulty of implementing such a complex innovation had been underestimated. But after the first four years the situation had changed significantly as teachers became skilled in handling the process, and found, for example, that their concerns that other teachers would play an advocacy role on behalf of their students largely unfounded. This trend was found to have continued in a survey conducted five years later.

Fairness :

Fairness emerged as a more complex issue than we had expected in the course of the evaluation. We focused on the fairness to students of the end of year review process. But attention should also have been given to the degree of similarity in the *conditions* under which students carried out the assessment tasks. There were indications that teachers varied in the extent to which they structured tasks for students, such as fieldwork, or expected them to take responsibility for planning their work. But monitoring this variation was beyond our resources. Case study evidence indicates that the group members gradually gravitated towards a more common level of assistance to students over time but differences still remain.

Teachers' confidence in the ability of their groups to moderate assessments for the Individual Research Projects fairly and reliably increased steadily, as indicated in Table 1. (From 80% to 95% agreeing, or strongly agreeing, with the statement).

Teacher confidence in the ability of their groups to carry out this part of the process was high from the beginning, mainly because the guidelines that geography teachers had developed for this task had a clear structure, as did the check-list which was used for assessing students' work. In addition, half way through the year, teachers were required to circulate outlines of their students' projects (including topics, aims, hypotheses, and method) to the other teachers in their CMG for approval as "geographic". As a result, by the end of the year the level of mutual understanding amongst the teachers about what each was doing was quite high.

Such clarity was not quite so possible for the Optional Unit. Here teachers could choose to teach one of two topics; "a geography of recreation", or "settlement patterns". And for assessment, as indicated above, there was more room for teacher discretion. They could choose three types of task out of six. In the first few years the assessment criteria for these tasks, such as fieldwork, were not so tightly defined as for the research project. Consequently, teachers were not so confident that they had compared students' work and grades across schools fairly and reliably. Over the ensuing years more explicit criteria were developed and the level of teacher confidence rose, as shown in Table 1, from 43% to 88%. But a problem still remains in the view of many teachers, and from next year guidelines for "common assessment tasks" will be introduced. These will reduce teacher discretion in selecting types of assessment somewhat, but improve the chances of groups to make meaningful comparisons in their moderation groups.

This change over time in teachers' views indicates the importance of regarding moderation as a complex innovation requiring a considerable period of time for learning, and un-learning during its implementation.

Surveys of students in the first two years indicated that they were in no doubt that the new system was fairer. Only 10% would have prefered a system based on an end of year examination alone. Contrary to the expectations of many, few detrimental effects on student-teacher relationships were reported by students as a result of their teachers taking on an assessor role. Where problems were reported, students' explanatory comments made it clear that they had more to do with the teacher's competence and style than with the new assessment procedures. Student interest and engagement in the course was very high. External examiners stated that they had noted an improvement in the quality of student answers in the external examination since moderation had been introduced. Some university academics volunteered the information that freshmen who had done the new course were noticeably more competent in carrying out projects requiring initiative and writing up research reports.

Reliability

It proved to be very difficult to gather definitive data about the measurement characteristics of assessment under consensus moderation. There was an uncanny capacity of groups of teachers to give almost identical marks for the same research project. Marked discrepancies were rare. It seemed at times as if teachers had the same inbuilt marking device, the way in which they could confidently declare "This is a 15 out of 20".

Teachers, to the surprise of some, produced as broad a spread of marks in their assessments as the external examination; that is, the capacity to discriminate was much the same. The number of changes made by moderation groups to teachers' assessments grew each year. Few changes were made in the first year. Teachers voiced concerns about some teachers' marks privately, but were tentative about pressing their views. In later years this diffidence disappeared.

Another way of getting at teachers' evaluation of the moderation process was by means of the questions set out in Table 2. This shows a similar trend to that of Table 1. Teachers have become steadily more satisfied that the grades that students finish up with are appropriate, although they are less confident about the grades given to work from other schools than that given to their own students. Again, they are less confident about the grades given for work in the option.

Since its introduction, school-based assessment and moderation has lacked credibility with some sections of the community. A strong faith remains in the fairness and objectivity of the external examination (even though on close inspection, the process of marking that actually goes on in the moderation groups is not really very different from that used by the external examiners, except that the former only look at samples of student work from a school.) At the present time the issue is very prominent because the State Labor Government is in the process of implementing a major reform of curriculum and assessment for the final two years of high school, and it has decided that a modified version of consensus moderation, called "panel verification": will now be used in <u>all</u> subjects. This has upset some sections of the universities and they have mounted a vigorous political campaign in opposition, threatening even to set up their own examinations to help them select students.

Table 3 provides responses to two questions designed to gauge teachers' views on this issue. The first asks for teachers to respond to a prediction from the Vice-Chancellor of the prestigious Melbourne University that the increased use of school-based assessment with moderation will make marks unreliable and lead, therefore, to universities placing more faith on a school's reputation and status than teachers' grades per se. The small number who agreed with this prediction (14%) mainly said they did so reluctantly because, no matter what evidence could be marshalled to the contrary, the powers that be in the universities would continue to believe that teachers moderated assessments could not be trusted. The dominant view was expressed in these terms by three teachers:

> From my experience with consensus moderation "trusted schools" have no advantage or higher standing over "other" schools - they all have equal standing. It is also up to individual teachers in the group to make sure that assessment is consistent - that is the whole point of it all, so marks should not end up being "unreliable". You're also in there for your kids, as are other teachers, and its also up to individuals to speak up - thereby limiting other schools being advantaged.

> The setting up of criteria for assessment at the beginning of the year greatly discourages any bias by "trusted" schools. In my experience, teachers are usually very honest about their students' abilities and the degree to which they have met our set criteria. Negotiation between teachers is usually conducted in a pleasant, calm and dignified manner until consensus is reached. Marks generally are reliable, with no particular advantage to be gained by being a teacher in a "good" school.
>
> If it is followed correctly under the Geography moderation guidelines this would not occur due to so many people actually looking at a particular student's work. It has inbuilt checks and balances that would hopefully make the system work.

Question 2 in Table 3 provided an opportunity to check a perception in some quarters that objectivity in consensus moderation is compromised by the strong advocacy played by some teachers and there is some cause for concern that nearly 20% of teachers agreed with the statement. One of them stated that

> I believe Consensus Moderation has a lot to offer, but I have had some terrible experiences where strong, loud and self-opinionated teachers who thought they knew all completely dominated and were so picky with the work (expected too much from Year 12). The convenor let these teachers run the show.
>
> I think some form of moderation is necessary as glaring errors over marking have been highlighted, but it should be used as a check, not a complete remarking process. Also the process is far too long and tiring for one day and more time should be put aside for it.

The majority of teachers, however, did not agree. These comments are representative.

> In my experience advantages gained in such a way are negligible. Our group works in a professional and fair manner for the vast majority of students involved in it. Marking has been reliable and consistent over the years to the benefit of all involved and to the process.
>
> The key is in establishing the assessment criteria at the start of the year. If they are clearly stated, then everyone will assess according to those. Particular schools have no advantages or disadvantages. The effectiveness of the convenor has a big impact on whether or not some individuals can "have their own way". In my experience, convenors have been very good in coping with dominant personalities.

There has been a lot of ironing out of wrinkles in the moderation process over the years, and some still remain. Several teachers in the last survey expressed concern that some teachers regularly ask if they can have the sample of their students work chosen by the convenor changed at the final meeting.

Workload

Initially there was concern that the workload was perceived by students to be greater than for other subjects. Teachers varied in their interpretation of how much work was required for the assessment tasks. Some over-zealous teachers apparently were determined that their

students would show up well in the process. There were some schools where every student reported that the workload was unreasonably high, whereas in most schools students thought the workload was fair. Over time the CMGs learned how to ensure that the workload was more consistent from school to school and that it was distributed evenly over the school year.

The increased workload for teachers however still remains an issue. Table 4 shows that a significant proportion of teachers have continued to feel that that the process relies too much on their time and goodwill to operate effectively, even though they believe in its merits in principle. Moderation is a considerable logistical exercise. Substitute teachers have been needed to replace teachers attending moderation meetings for two days per year, but most teachers have always argued that at least another day should be provided for the end of year meeting if moderation is to be carried out thoroughly. The government intends to get around this problem by declaring extra "pupil-free days" on the school calendar when moderation comes in for all subjects.

Professional development and accountability

Perhaps most importantly moderation represented a widespread and successful exercise in professional development and self esteem. The best way to convey a sense of that development would be through case studies of the work of individual groups over time. Space precludes providing those here, but Table 3 provides a summary of surveys of teachers' views on the contribution which moderation made. For example, in 1989, 89% of teachers agreed or strongly agreed that the process had added significantly to the skills they possessed for assessing student learning. 85% thought that it had contributed significantly to their ability to evaluate and improve their teaching, and 82% thought the process had meant that their access to useful ideas for teaching had been increased significantly. Most felt that the quality of learning of their students had been enhanced (76%) and that there had been a flow on effect to their teaching in other classes (68%). For most of these questions, the proportion of teachers who responded positively increased with their experience of moderation. 85% agreed that their involvement had meant that they had become more accountable to professional colleagues.

Inexperienced teachers were particularly favourable about the support they had gained, even though at first the prospect of facing up to a group of 10-12 experienced teachers was very daunting.

Principals had been surveyed in 1983, after three years of moderation, and ninety percent thought that the process had made a positive contribution to the professional development of the teachers from their school who were involved. 87% thought that it had enhanced the quality of the teaching methods that they used.

At the same time moderation provided opportunities for leadership as a convenor. There has been no difficulty in recruiting new teachers to replace retiring convenors, even though they only receive a token $500 for many hours of work, and considerable demands upon their private time. Moderation by consensus is crucially dependent on the social skills and personal qualities of the people who act as convenors for each moderation group. No one should embark on such an innovation without a guarantee that a suitable supply of people is available. They need to be experts with respect to the course, and the regulations which cover the moderation process, which are extensive. They need to be trained in verbal and problem-solving skills to facilitate decision-making based on consensus, and able to mould a number of teachers into a congenial, task-oriented group. Interviews with convenors indicated that they had gained satisfaction from undertaking the new responsibility, but it was demanding.

The meaning of moderation

It has been interesting to observe the gradual development of a shared and richer understanding of the concept of moderation over the past ten years (c.f. Fullan, 1982) For many participants in the early years of the process, moderation was what happened at the end of year meeting when teachers looked over each other's student's work and assessments. Gradually, the view has emerged that moderation also includes ensuring comparability in the interpretation of the objectives, the tasks set, the conditions under which students do the tasks, the assessment techniques used, and the teachers' views of standards. In this view, moderation includes the context as well as the outcomes, and it appears that if the contextual aspects are monitored during the year, consensus moderation groups have fewer difficulties in agreeing on final assessment s at the end of the year.

Consensus moderation can be regarded as a process through which trust is built up amongst a group of teachers. This trust is based on a mutual understanding of the activities each is using to assess the degree of achievement of the objectives of the course, and why. It also includes confidence that the conditions under which these activities are taking place are comparable for students from different schools.

During the year, when members of a consensus moderation group circulate amongst themselves their proposals for fieldwork, practical work, reasearch projects, and so on, validity has been promoted in that a check has been carried out on the extent to which these tasks are relevant ot the objectives of the course. Fairness has been promoted to the degree that the group has ensured comparability in the conditions under which students from different schools demonstrate what they can do, and teachers' accountability has been increased through the sharing and review of each other's activities.

Final comments

The findings of this study can be related to several aspects of the current educational reform debate in the U.S.. Consensus moderation is an example of an area responsibility which enhances professionalism and professional development. It is a vehicle which provides many opportunities for teacher leadership, collegiality and the sharing of professional knowledge about curriculum and assessment. It also provides more opportunities for genuine participation in decision-making about matters close to teachers' workplace concerns, such as curriculum and assessment and the setting of standards.

Teachers' grades really matter under this system, and as a result the process provides the kind of incentive for student effort that the president of the American Federation of Teachers was calling for in his recent address to the AFT QuEST 1989 conference on "Restructuring Schools".

The process supports teachers by giving weight to their assessments; and it reinforces and clarifies the links between the modes of assessment teachers choose and their curriculum objectives. These links are weakened when the results of standardised aptitude and achievement tests are given greater weight than teachers' own assessments of their students' learning. Moderation reinforces teachers' engagement with their work. They feel directly responsible for their assessments. Teachers' work is much more open to their professional peers. When members of moderation groups share their proposals for fieldwork, practical work, and student research projects for review and approval, not only is the validity and fairness of their assessments improved; a powerful, and, to most teachers, justifiable, mode of professional accountability has been introduced.

References

Calfee, R. C. and Hiebert, E. (1988). *The Teacher's Role In Using Assessment To Improve Learning*. Center for Educational Research at Stanford, Stanford University, Stanford.

Carnegie Forum On Education And The Economy. (1986). *A Nation Prepared: Teachers for the 21st Century*. New York: The Carnegie Forum

Darling-Hammond, L. (1989). Accountability fo Professional Practice.Teachers College Record, 91(1), 59-80.

Fullan, M. (1982). *The Meaning of Educational Change*. New York: Teachers College Press.

Johnson, S. M. (1989) .Schoolwork and Its Reform. In Jane Hannaway and Robert Crowson, *The Politics of Reforming School Administration.* New York: The Falmer Press

Stake, R. (1989). *Effects of Changes in Assessment Policy.* JAI Press

Stake, R. (1988). *The effects of reform in assessment in the USA*. Paper presented at the annual meeting of the British Educational Research Association at the University of East Anglia, September, 1988

TABLE 1

Teachers' attitudes to the moderation process

1.(a) After my consensus moderation group had dealt with the Individual Research Project at the November meeting, I had a feeling of confidence and trust in my group's ability to assess students' work fairly and reliably.

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1981	19	61	10	1	8
1982	19	57	15	7	1
1983	39	53	6	0	2
1984	33	59	4	1	3
1989	35	60	4	1	0

(b) After my consensus moderation group had dealt with the Optional Unit at the November meeting, I had a feeling of confidence and trust in my group's ability to assess students' work fairly and reliably.

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1981	4	39	31	17	10
1982	10	52	24	7	6
1983	20	55	14	4	7
1984	17	59	15	2	7
1989	23	65	6	1	4

TABLE 2

Teachers' attitudes to the moderation of school-based assessments

2. After the November meeting of my C.M.G., I was satisfied about the appropriateness of the Independent Research Project marks allocated to:

(i) my students

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1981	24	58	13	1	4
1982	27	58	10	4	0
1983	44	50	5	0	1
1984	30	64	4	1	1
1989	33	61	2	0	4

(ii) students from other schools

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1982	21	55	16	3	5
1983	31	58	7	0	3
1984	21	67	6	1	5
1989	15	73	7	0	5

3. After the November meeting of my C.M.G., I was satisfied about the appropriateness of the Optional Unit marks allocated to:

(i) my students

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1981	13	58	8	6	15
1982	19	72	4	0	3
1983	28	62	6	1	3
1984	20	67	7	1	5
1989	26	67	5	0	2

(ii) students from other schools

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1982	9	60	18	3	10
1983	15	61	11	2	10
1984	13	64	11	2	10
1989	15	66	12	0	7

TABLE 3

1. As you will know already, moderation, or "verification" as it will be called under the new Victorian Curriculum and Assessment Board arrangements for student assessment, has been the subject of public debate. According to "The Sunday Age" (Feb. 4, 1990), for example, some people argue that

> the new assessment procedures - in which three of the four Common Assessment Tasks will be assessed internally and then moderated against district schools - will make marks unreliable and lead to "trusted schools" gaining advantages for their students. (p. 9)

Given your experience with consensus moderation in Geography, would you agree with this prediction?

	Yes	No
Government schools	12	88
Catholic schools	12	88
Independent schools	21	79
Total	14	86

2. The objectivity of moderation is highly questionable because some strongly opinionated teachers sway the assessments of group members on behalf of their own students

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1989	2	17	52	23	6

TABLE 4

The benefits of consensus moderation are not commensurate with the time and effort required from both teachers and students

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1981	15	21	32	13	19
1982	12	30	22	24	12
1983	14	21	39	19	8
1984	13	15	40	21	8
1989	7	19	41	16	17

I would recommend consensus moderation to others who were considering adopting it.

	Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
1984	22	48	9	7	14
1989	25	53	7	4	11

TABLE 5

Effects of Moderation on the Professional Development of Teachers

The consensus moderation process has:

		Strongly Agree	Agree	Disagree	Strongly Disagree	Undecided
Contributed significantly to the skills I possess for assessing student learning	1983	29	47	13	1	10
	1984	34	53	6	1	6
	1989	33	56	5	1	5
Contributed significantly to my ability to evaluate and improve my own teaching	1983	22	53	14	0	11
	1984	28	54	8	2	7
	1989	20	65	8	1	6
Contributed significantly to the quality of the learning of my students	1983	18	40	23	2	16
	1984	19	49	12	2	18
	1989	14	62	13	2	9
Meant that my access to useful ideas for teaching has been increased significantly.	1983	23	51	17	1	8
	1984	24	61	10	1	4
	1989	27	55	10	1	7
Influenced beneficially the teaching methods I use in other (non Year 12) classes.	1983	11	43	30	5	12
	1984	12	56	16	2	14
Increased the amount of independence in learning displayed by my students.	1984	15	45	18	4	18
	1989	4	55	23	2	17
Meant that I have become more accountable to professional colleagues	1989	25	60	8	2	5